DOI:
10.1136/gutjnl-2014-308896

Document Version
Peer reviewed version

Crohn's disease siblings exhibit a biologically relevant dysbiosis in mucosal microbial metacommunities

Authors

Charlotte Hedin[1,5a,] Christopher J. van der Gast[2a], Geraint B. Rogers[3], Leah Cuthbertson[2], Sara McCartney[4], Andrew J. Stagg[5], James O. Lindsay[6,7], Kevin Whelan[1]

[1] King's College London, Faculty of Life Sciences & Medicine, Diabetes and Nutritional Sciences Division, London, United Kingdom

[2] NERC Centre for Ecology & Hydrology, Wallingford, Oxfordshire, United Kingdom

[3] South Australian Health and Medical Research Institute, Infection and Immunity Theme, Flinders University, Adelaide, Australia

[4] University College London, Centre for Gastroenterology and Nutrition, London, United Kingdom

[5] Blizard Institute, Queen Mary University of London, Centre for Immunology and Infectious Disease, London, United Kingdom

[6] Blizard Institute, Queen Mary University of London, Centre for Digestive Diseases, London, United Kingdom

[7] Barts Health NHS Trust, Department of Gastroenterology, London, United Kingdom

[a] These authors contributed equally to this work

Corresponding author

Professor Kevin Whelan

King's College London, Faculty of Life Sciences & Medicine, Diabetes and Nutritional Sciences Division, Franklin Wilkins Building, 150 Stamford Street, London, SE1 9NH, United Kingdom

Tel: +44 (0)20 78 48 38 58

Email: kevin.whelan@kcl.ac.uk

Word count: 4162

Number of figures: 4

Number of tables: 3

Number of references: 39

Number of supplementary files for online only publication: 3

Data deposition

The sequence data reported in this paper have been deposited in the NCBI Short Read Archive database (Accession number SRP045959)

Competing interest statement: None declared

Author contributions

CRH: study concept and design; obtained funding; recruitment of participants and acquisition of data; analysis and interpretation of data; statistical analysis; drafting of the manuscript; critical revision of the manuscript for important intellectual content. CvdG: Analysis and interpretation of data; statistical analysis; drafting of the manuscript; critical revision of the manuscript for important intellectual content. GBR: DNA extraction; analysis and interpretation of data; statistical analysis; critical revision of the manuscript for important intellectual content. LC: analysis and interpretation of data; statistical analysis. SM: assistance with recruitment of participants; critical revision of the manuscript for important intellectual content. AJS, JOL and KW: study concept and design; obtained funding; analysis and interpretation of data; critical revision of the manuscript for important intellectual content; study supervision.

ABSTRACT

Objective

Siblings of patients with Crohn's disease (CD) have elevated risk of developing CD and display aspects of disease phenotype, including faecal dysbiosis. Whether the mucosal microbiota is disrupted in these at-risk individuals is unknown. Objective: To determine the existence of mucosal dysbiosis in siblings of CD patients using 454 pyrosequencing and to comprehensively characterise, and determine the influence of genotypic and phenotypic factors, on that dysbiosis.

Design

Rectal biopsy DNA was extracted from 21 patients with quiescent CD, 17 of their healthy siblings and 19 unrelated healthy controls. Mucosal microbiota was analysed by 16S rRNA gene pyrosequencing and were classified into core and rare species. Genotypic risk was determined using Illumina Immuno BeadChip, faecal calprotectin by ELISA and blood T-cell phenotype by flow cytometry.

Results

Core microbiota of both CD patients and healthy siblings were significantly less diverse than controls. Metacommunity profiling (Bray-Curtis (S_{BC}) index) showed the sibling core microbial composition to be more similar to CD (S_{BC}=0.70) than to HC, whereas the sibling rare microbiota was more similar to HC (S_{BC}=0.42). *Faecalibacterium prausnitzii* contributed most to core metacommunity dissimilarity both between siblings and controls, and between patients and controls. Phenotype/genotype markers of CD-risk significantly influenced microbiota variation between and within groups, of which genotype had the largest effect.

Conclusion

Individuals with elevated CD-risk display mucosal dysbiosis characterised by reduced diversity of core microbiota and lower abundance of *F. prausnitzii*. This dysbiosis in healthy people at-risk of CD implicates microbiological processes in CD pathogenesis.

SUMMARY BOX

What is already known about this subject:

- Patients with CD have mucosal dysbiosis, including reduced abundance of *Faecalibacterium prausnitzii*
- Low mucosal *Faecalibacterium prausnitzii* predicts relapse after surgery in CD patients
- Healthy siblings of CD patients have increased risk of developing CD and have altered abundance of key species in the gut lumen

What are the new findings:

- There is a distinct dysbiosis in the mucosal microbiota of healthy siblings of CD patients
- The sibling dysbiosis comprises a fundamental distortion of microbial community composition, most notably reduced diversity of core microbiota and low abundance of mucosal *Faecalibacterium prausnitzii*
- Mucosal microbiota disruption is not merely a consequence of the inflammation in CD but is present at healthy individuals at risk of CD

How might it impact on clinical practice in the foreseeable future?

- Identification of this at risk dysbiosis signals pathways in CD pathogenesis and raises the possibility of CD risk identification and CD risk intervention

INTRODUCTION

Disruption of gut microbiota (dysbiosis) is an established feature of inflammatory bowel disease (IBD). The dysbiosis in Crohn's disease (CD) has been well described and includes reduced microbial diversity, reduced abundance of Firmicutes particularly *Faecalibacterium prausnitzii*, reduced abundance of Bifidobacteria, increased γ-proteobacteria and disturbances in Bacteroides populations.[1] The involvement of several CD susceptibility genes in the recognition and handling of bacteria (e.g., NOD2, ATG16L1, IRGM) reinforces the position of the gut microbiota at the centre of IBD pathogenesis. [2, 3, 4]

Whether the CD dysbiosis is involved with pathogenesis is uncertain. The dependence on the presence of gut microbiota for the development of inflammation in animal models[5] as well as CD patients,[6] and the association between reduced mucosal *F. prausnitzii* and post-operative relapse[7] implies a pathogenic role. Conversely, the lack of therapeutic benefit of manipulating the microbiota,[8,9] suggests that dysbiosis in CD may not drive inflammation, but rather is consequent to established disease, reflecting for example, the differential survival of various species in an inflamed environment. Moreover, attempts to identify aspects of the CD dysbiosis that were present at disease initiation, which therefore potentially have a role in pathogenesis may be obfuscated by both the mature disease phenotype of the patients studied and the effect of the medical, surgical and patient-initiated attempts to treat and control symptoms.

Siblings of CD patients have a relative risk (RR) of developing CD of up to 35 times that of the general population.[10] This risk is partly genetic, but is also driven by non-genetic factors many of which they share with their CD-affected sibling.[10,11] Several of these non-genetic risk factors, such as mode of delivery, breast feeding, maternal inoculum, home environment and weaning diet,[12] potentially impact gut microbial acquisition and development. It follows that any aspect of the CD dysbiosis which is also present in a healthy sibling cannot be disrupted as a consequence of disease, and rather may be implicated in processes driving CD pathogenesis.[12]

Attempts have been made to determine whether aspects of the CD phenotype are present in patients' unaffected relatives. These have assessed dysbiosis[13] and other features of the CD phenotype such as raised faecal calprotectin (FC), increased intestinal permeability (IP) and the presence of anti-microbial antibodies.[12] Using PCR probes selected to detect dominant species that comprise the dysbiosis in CD, we have previously indicated that a faecal dysbiosis exists in healthy siblings of CD patients characterised by reduced faecal Firmicutes including *F. prausnitzii*.[14] Moreover, we previously demonstrated in siblings that a combination of luminal dysbiosis, raised FC, reduced abundance of circulating naïve T-cells, disturbances in their expression of gut-homing β7 integrin and at-risk genotype could be combined to create a multidimensional risk phenotype, which significantly distinguished healthy siblings of CD patients from healthy, unrelated controls.[14]

It has been speculated that mucosal microbiota are of greater significance in CD pathogenesis than luminal microbiota given their closer spatial relationship to the gut immune system. Yet, studies comparing mucosal microbiota in CD patients, their families and healthy controls are rare due to the invasiveness of procedures required to obtain mucosal samples from otherwise healthy individuals. However, the potential rewards of obtaining such samples have been amplified by recent advances in the analysis of large, diverse and complex microbial communities. Pyrosequencing technology and meta-community profiling enables sampling depth permitting detection not only of dominant microbial community members but also low-abundance (rare) taxa.[15, 16] The capacity to characterise core and rare microbial communities separately may reveal microbial features associated with disease not otherwise readily apparent. Furthermore, 16S rRNA gene pyrosequencing and other next-generation technologies have demonstrated that microbial diversity can be orders of magnitude higher than previously appreciated.[16] Measuring diversity may be significant as healthy gut microbiota high diversity compared with microbial populations in other human body habitats.[18] Moreover, gut microbial diversity is consistently described as reduced both in CD,[1] and other human diseases including obesity,[18, 19] colorectal cancer,[20] eczema,[21] and in addition has been linked with smoking.[22]

Therefore, we sought to use 454 pyrosequencing and metacommunity analysis to comprehensively characterise the structure and composition of the mucosal microbial community in an at-risk group of CD siblings compared with CD patients and healthy controls.

MATERIALS AND METHODS

Patients with inactive CD (Crohn's Disease Activity Index (CDAI) <150 and C-reactive protein (CRP) ≤5mg/L, and their healthy siblings (both 16-35 years) were recruited from clinics at Barts Health NHS Trust and University College Hospitals NHS Foundation Trust (London, UK). Patients required a confirmed diagnosis of CD for >3months. All healthy siblings who volunteered and did not meet exclusion criteria (detailed in supplementary Table S1) were included, to limit bias in the selection of siblings with specific characteristics. Healthy controls were recruited by email sent to staff and students at King's College London (London, UK), during the same period. Participants were informed that involvement in the study did not constitute screening for disease and that detection of clinical disease in any sibling or control would lead to exclusion from the study.

Only participants consenting to rectoscopy and providing analysable biopsies were included. All participants provided written, informed consent. Ethical approval was provided by Bromley Local Research Ethics Committee (reference 07/H0805/46).

At screening, demographics, medical and drug exposure history, physical examination, CRP, inclusion and exclusion criteria were assessed. Instructions regarding avoidance of prebiotics/probiotics for 4 weeks (to prevent impact on microbiota), non-steroidal anti-inflammatory drugs for 1 week and alcohol for 24h before the study (to prevent impact on IP) were provided. Blood samples were taken for routine haematology/biochemistry, T-cell analysis and genotyping. Participants completed a 5h urine collection for measurement of IP and underwent flexible rectoscopy without bowel cleansing. Biopsies from non-inflamed rectum were snap frozen, and stored at -80^{o}C before processing for histological and microbiological analyses. Stool was obtained and stored at -20^{o}C before processing for FC quantification.

Faecal calprotectin

FC extraction and ELISA analysis (Calpro AS, Lysaker, Norway) were carried out according to manufacturer's instructions using duplicate appropriately diluted samples. FC concentration (μg/g) was determined relative to standard curves.

Peripheral blood T-cell flow cytometry

Whole blood, collected in lithium-heparin Vacutainer tubes (BD Bioscience), was stored at room temperature for ≤4 h before labelling with fluorescently conjugated monoclonal antibodies to detect CD3 T-cells, naïve ($CD45RA^+$) and memory ($CD45RA^-$) subsets of CD4 and CD8 T-cells. Integrin α4β7 expression was assessed by labelling with anti-β7 (see supplementary methods for antibodies used). Data were acquired using a LSRII 4-colour flow cytometer (BD Bioscience) and collected using fluorescence-activated cell sorting Diva software V.4.1.2 (BD Bioscience) using Flow-Count fluorospheres (Beckman Coulter) for absolute quantitation. Colour compensation was performed offline using Winlist V.6.0 (Verity Software House).

Genotyping

Human DNA was extracted from whole blood using the phenol chloroform-isoamyl alcohol method. Genotyping was performed using the Illumina Infinium Immunochip.[2, 23] To increase detection of NOD2 mutations and capture the enhanced risk of NOD2 compound heterozygosity, three NOD2 mutations (rs2066845/G908R, rs2066844/R702W and rs5743293/3020insC) were individually assessed. Cumulative genotype relative risk (GRR) for each participant was therefore calculated across 72 CD-risk loci. A population distribution model of CD-risk was generated using the REGENT R program[24] and previously published odds ratios.[2] Participants were categorised into reduced, average, elevated or high genotype risk with reference to this model.[25]

Intestinal permeability

IP was measured using lactulose-rhamnose tests as previously described.[14]

Gut mucosal microbiota

Biopsy DNA extraction was carried out using a phenol/chloroform based method, as described previously.[26] A detailed extraction protocol is provided in the supplementary methods. DNA extracts were quantified using the Picodrop Microlitre Spectrophotometer (GRI, Braintree, UK). Negative controls (sterile water), were included in the DNA extraction and PCR amplification steps.

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed as described previously using Gray28F 5'-TTTGATCNTGGCTCAG-3' and Gray519r 5'-GTNTTACNGCGGCKGCTG-3'.[27] Detailed protocols for 16S rRNA gene sequencing and sequence data processing are provided in the supplementary methods.

To assign bacterial identities to 16S rRNA gene sequences, sequence data were de-noised, assembled into OTU clusters at 97% identity, and queried using a distributed .NEt algorithm that utilises Blastn+ (KrakenBLAST, www.krakenblast.com) against a database of high quality 16S rRNA gene bacterial sequences. Using a .NET and C# analysis pipeline the resulting BLASTn+ outputs were compiled, data reduction analysis performed, and sequence identity classification carried out, as described previously.[28]

Statistical analyses

Bacterial species within each metacommunity were partitioned into common and rare groups using a modification of a previously described method.[15] Three complementary measurements of diversity were used to compare microbial diversity between samples, as previously described: species richness (S*, the total number of species), Shannon-Wiener (H', a metric accounting for both number and relative abundance of species), and Simpson's (1-D, a measure of the probability that two species randomly selected from a sample will differ).[15, 26] To avoid potential bias due to varying sequences per sample, all measures were calculated using randomised re-sampling to a uniform number of sequence reads per sample.[26] Mean diversity measures were calculated from the re-sampling of the reads from each specimen to the lowest number of sequence reads among all specimens for 1000 iterations. Diversity analysis was performed in R.[29] Two sample t-tests, regression analysis, coefficients of determination (r2),

residuals and significance (P) were calculated using Minitab software (version 16, Minitab, University Park, PA, USA). Canonical correspondence analysis (CCA), analysis of similarity (ANOSIM), similarity of percentages (SIMPER) analysis were performed using the PAST (Palaeontological Statistics, version 3.01) program available from the University of Oslo website link (http://folk.uio.no/ohammer/ past) run by Øyvind Hammer. The Bray-Curtis quantitative index of similarity was used as the underpinning community similarity measure for CCA, ANOSIM, and SIMPER tests.

RESULTS

Demographic and disease characteristics of the 21 patients with quiescent CD, 17 of their healthy siblings, and 19 unrelated healthy controls that were included are summarised in Table 1. At the time of the study only one patient was cohabiting with one of the included siblings. GRR, FC, faecal Firmicute abundance and circulating T-cell characteristics were all significantly different in both CD patients and healthy siblings compared with healthy controls as previously published,[14] and as summarised in Table 1.

Table 1 Summary of demographic variables in patients, siblings and controls as well as clinical characteristics in patients. The features of the at-risk phenotype that have previously been delineated in this cohort are also displayed.

		Patients (n=21)	Siblings (n=17)	Controls (n=19)	*P*-value
Mean age years, (SD)		27.7 (6.6)	25.5 (4.5)	27.7 (5.8)	0.783*
Males, n (%)		13 (62)	11 (65)	9 (47)	0.515†
Body Mass Index, kg/m^2 (SD)		24.5 (5.0)	24.5 (3.6)	23.9 (3.4)	0.870*
Ethnicity n (%)	White British	17 (81)	15 (88)	17 (90)	0.469†
	Asian/Asian British	3 (14)	1 (6)	0 (0)	
	Black British or mixed black/white	1 (5)	1 (6)	2 (11)	
Smoking n (%)	Never	14 (67)	10 (59)	12 (63)	0.830†
	Current	4 (19)	5 (29)	3 (16)	
	Previous	3 (14)	2 (12)	4 (21)	
Age at diagnosis, n (%)	Below 16 years	7 (33)			
	16-40 years	14 (67)			
Disease location, n (%)	Ileal	7 (33)			
	Colonic	5 (24)			
	Ileocolonic	9 (43)			
Concomitant upper GI disease, n (%)		1 (5)			
Disease behaviour, n (%)	Non-stricturing, non-penetrating	11 (52)			
	Stricturing	5 (24)			
	Penetrating	5 (24)			
Perianal disease, n (%)		4 (19)			
Current 5-ASA n (%)		11 (52)			
Current immuno-suppressant, n (%)	Azathioprine	7 (33)			
	Mercaptopurine	2 (10)			
	Methotrexate	1 (5)			
Ileocaecal resection / right hemicolectomy, n (%)		9 (43)			
Isolated small bowel resection, n (%)		1 (5)			
Genotype relative risk,‖ n (%)	High	3 (14)	1 (6)	0 (0)	0.175†
	Elevated	2 (10)	3 (18)	0 (0)	
	Average	10 (48)	8 (47)	8 (42)	

Reduced	6 (29)	5 (29)	11 (58)	
Fecal calprotectin, m/g (IQR)	281 (144-855)	30 (13-83)	13 (7-33)	<0.001‡
Faecal *F. prausnitzii*, % (IQR)	0.1 (0.0-2.9)	3.7 (0.4-7.1)	5.2 (2.3-7.2)	0.001‡
T-cells with memory phenotype, % (IQR)	73 (63-82)	74 (67-83)	65 (54-70)	0.011‡
Naïve $CD4^+$ T-cells, cells/ ml (IQR)	194,132 (71,053-341,156)	198,220 (128,550-296,351)	380,256 (279,118-564,861)	<0.001‡
Naïve $CD4^+$ T-cells expressing β7 integrin, % (IQR)	76 (63-85)	74 (61-83)	52 (32-71)	0.003‡
Intestinal permeability: urinary lactulose-rhamnose ratio, (IQR)§	0.061 (0.033-0.111)	0.034 (0.024-0.056)	0.038 (0.025-0.050)	0.081‡

* One-way ANOVA
† Chi-squared test
‡ Kruskall-Wallis test
§ Data from 20 patients, 17 siblings and 16 controls contributed to the intestinal permeability analysis
ǁ Cumulative genotype relative risk (GRR) for each participant was calculated across 72 CD-risk loci (detected using the Illumina Infinium Immunochip). Participants were categorised into reduced, average, elevated or high genotype risk with reference to a population distribution model of CD-risk, previously described.[14]

A total of 180,696 bacterial sequence reads (mean per sample 3235 ± SD 205), identifying 160 genera and 351 distinct operational taxonomic units (OTUs) classified to species level (Table S2), were generated from all samples combined. The numbers of bacterial sequence reads per sample were similar among the three cohorts (mean ± SD): CD, 3296 ± 258 (n =21); siblings, 3190 ± 423 (n =17); and healthy, 3210 ± 393 (n =19).

Species abundance was directly correlated with distribution

We have previously established that the categorisation of human microbiota into core and rare species revealed important aspects of metacommunity species-abundance distributions that would be neglected without such a distinction.[15] A coherent metacommunity could be expected to exhibit a direct relationship between prevalence and abundance of individual species within the constituent communities. Consistent with this prediction, the abundance of species in each study group significantly correlated with the number of individual sample communities those species occupied (CD (R^2 =0.62, $F_{1,227}$ =366.9, P < 0.0001); siblings (R^2 =0.71, $F_{1,259}$ =590.1, P < 0.0001); and healthy controls (R^2 =0.68, $F_{1,\ 258}$ =552.6, P < 0.0001)), (Fig. 1).

In CD patients a lower proportion of the mucosal microbiota were core species

Individual species in each cohort metacommunity were then classified as core or rare based on their falling within or outside the upper quartile of subject occupancy, respectively (Fig. 1). Of the 229 species that comprised the CD metacommunity, only 7 were core and 222 were rare species. The healthy siblings metacommunity (261 species) comprised 18 core and 243 rare species, and the healthy controls metacommunity (260 species) comprised 25 and 235 species, respectively. In addition, the core species within each cohort metacommunity accounted for 44.7% ± 4.8% (CD), 67.6%,± 5.5% (healthy siblings) and 67.4% ± 4.6 (healthy controls) of the mean (± SD) relative abundance. The mean relative abundances in the CD core microbiota were significantly lower than the healthy siblings and healthy controls (P < 0.0001 in both cases), but were not different between the siblings and healthy controls (P =0.907).

Microbial diversity was lower in both siblings and patients compared with controls

The mean microbial diversity of subject communities for each cohort was compared using three indices of diversity (Fig. 2). Diversity was compared between the three cohorts for the whole microbiota, as well as core and rare species groups (Fig. 2). These analyses revealed the siblings' whole and core microbiota to be significantly more diverse than the CD cohort, but the sibling core microbiota was significantly less diverse than the healthy core microbiota. No significant difference in diversity was observed between the whole microbiota between the siblings and healthy cohorts, emphasising the advantage of analysing core and rare populations separately. In addition, the CD rare microbiota was significantly less diverse than the other two rare species cohorts, which in turn were not significantly different from each other. All of these observations were underpinned by all three measures of diversity in each instance (Fig. 2).

Interestingly, within the CD population, diversity of the whole microbiota was lower in the nine patients with an ileocaecal resection / right hemicolectomy compared with the 11 patients without these operations (as shown by Richness $P<0.0001$; Shannon-Wiener $P=0.046$; but not Simpson's $P=0.768$). This was largely driven by lower diversity of rare taxa (as shown by Richness $P<0.0001$; Shannon-Wiener $P=0.019$; but not Simpson's $P=0.159$) rather than core taxa (Richness $P=0.523$; Simpson's $P=0.612$; Shannon-Wiener $P=0.824$).

Significant divergence in whole and core microbial composition between CD patients and healthy controls, but not between CD patients and healthy siblings

The distribution of the microbiota within the three cohorts was determined by direct ordination using Bray-Curtis similarity measures. Using Analysis of Similarities (ANOSIM) tests, the CD and healthy whole and core microbiota were significantly divergent from each other. However, the whole and core microbiota of siblings were not significantly divergent from either that of the CD or healthy controls (Fig. 3). In all instances rare microbiota were significantly divergent between cohorts, including between siblings and healthy controls.

Lower *Faecalibacterium prausnitzii* made the greatest contribution to the dissimilarity in microbiota between both healthy siblings and healthy controls and between CD and healthy controls

Given the involvement of core species in differences of relative abundance, diversity and microbiota composition, the contribution of individual taxa to the dissimilarity between core microbiota was assessed by Similarities of Percentages (SIMPER) analyses (Table 2). Both *F. prausnitzii* and *Escherichia fergusonii* contributed the most to the dissimilarity between all cohorts. As a proportion of core species *F. prausnitzii* had a higher relative abundance in the healthy controls (30.9%) than both the CD (22.4%) and siblings (24.2%). Conversely, *E. fergusonii* was more abundant in the CD cohort (21.4%) than in siblings (9.7%) and healthy controls (4.1%).

Table 2 Similarity of Percentages (SIMPER) analysis of microbial community dissimilarity (Bray-Curtis) between core species groups for (A) CD and siblings, (B) healthy and siblings, and (C) CD and healthy cohorts. Given is mean % abundance of sequences for core species only across the samples each was observed to occupy and the average dissimilarity between samples (overall mean (A) =73.4% and (B) =55.0%, (C) =73.0%). Percentage contribution is the mean contribution divided by mean dissimilarity across samples. The list of species is not exhaustive so cumulative % value does not sum to 100%. Species level identities of detected taxa are reported here. However, given the length of the ribosomal sequences analysed, these identities should be considered putative.

A	Crohn's	Siblings		
Name	mean abundance	mean abundance	% Contribution	Cumulative %
Faecalibacterium prausnitzii	22.4	24.2	20.7	20.7
Escherichia fergusonii	21.4	9.7	15.9	36.6
Shigella flexneri	13.6	7.2	10.7	47.3
Ruminococcus gnavus	13.1	5.2	8.9	56.2
Bacteroides vulgatus	13.2	7.6	7.8	64.0
Eubacterium rectale	9.8	6.4	6.6	70.6
Oscillospira guilliermondii	0	8.0	5.9	76.5
Escherichia coli	6.5	0	4.5	81.0
Sutterella wadsworthensis	0	6.0	4.5	85.5
Bacteroides dorei	0	5.7	4.2	89.6
Roseburia faecis	0	4.0	2.9	92.6
B	Healthy	Siblings		
Name	mean abundance	mean abundance	% Contribution	Cumulative %
Faecalibacterium prausnitzii	30.9	24.2	18.9	18.9
Escherichia fergusonii	4.1	9.7	10.6	29.5
Sutterella wadsworthensis	8.7	6.0	9.4	38.9
Shigella flexneri	3.6	7.2	8.4	47.3
Bacteroides vulgatus	8.0	7.6	8.4	55.7
Eubacterium rectale	9.9	6.4	7.0	62.8
Oscillospira guilliermondii	8.5	8.0	7.0	69.8
Bacteroides dorei	0	5.7	5.4	75.2
Ruminococcus gnavus	4.1	5.2	4.1	79.3
Bacteroides uniformis	2.9	2.0	3.1	82.4
Roseburia faecis	2.4	4.0	3.0	85.4
Coprococcus eutactus	2.3	0	2.2	87.7
Shigella dysenteriae	2.1	0	2.1	89.8
Blautia producta	2.0	1.8	1.8	91.6
C	Crohn's	Healthy		
Name	mean abundance	mean abundance	% Contribution	Cumulative %
Faecalibacterium prausnitzii	22.4	30.9	22.4	22.4
Escherichia fergusonii	21.4	4.1	14.4	36.7
Shigella flexneri	13.6	3.6	9.3	46.0
Ruminococcus gnavus	13.1	4.1	8.6	54.6
Bacteroides vulgatus	13.2	8.0	8.0	62.7
Eubacterium rectale	9.8	9.9	7.4	70.1
Sutterella wadsworthensis	0	8.7	6.4	76.5
Oscillospira guilliermondii	0	8.5	6.2	82.8
Escherichia coli	6.5	0	4.4	87.2
Bacteroides uniformis	0	2.9	2.1	89.3
Roseburia faecis	0	2.4	1.8	91.1

Genotype and phenotypic features associated with CD and CD-risk significantly explained microbiota variation

Canonical correspondence analysis (CCA) was used to relate the variability in the distribution of microbiota between cohorts to clinical and demographic variables (Table 3 and Fig. 4). Variables that significantly explained variation in mucosal microbiota were determined with forward selection (999 Monte Carlo permutations; $P < 0.05$) and used in CCA. Based on the direct ordination approach, the microbiota between cohorts was significantly influenced by factors listed in Table 3. The same analytical approach was used to assess the extent to which variance in the microbiota distribution within cohorts could be accounted for by variation in measures of clinical and demographic factors, (Table 3). GRR was the most significant factor in explaining variance between the three cohorts, but also within each cohort. FC was also significant in explaining variance between cohorts, particularly in the core microbiota. However, in the within-group analyses FC was significant in explaining microbial variance in patients and siblings but not in controls. Blood T-cell factors explained a higher proportion of variance in siblings and controls than in patients. Conversely, age significantly associated with variance in controls but not in patients or siblings.

316 **Table 3 Canonical correspondence analyses for determination of percent variation in the whole, core, and rare microbiota between and within the three**

317 **subject cohorts by clinical variables significant at the P < 0.05 level**. * Ileal/Colonic involvement in CD patients used as a factor for the corresponding siblings. n/a

318 denotes not applicable for between cohort or within healthy cohort analyses.

	Between Cohorts			Within Crohn's			Within Siblings			Within Healthy		
Variable	**Whole**	**Core**	**Rare**	**Whole**	**Core**	**Rare**	**Whole**	**Core**	**Rare**	**Whole**	**Core**	**Rare**
Age	-	-	-	-	-	-	-	-	-	5.37	6.08	5.08
Blood concentration of naïve $CD4^+$ T-cells (cells /ml)	-	2.8	-	4.15	3.69	4.54	7.34	8.84	8.30	-	-	-
Calprotectin	1.7	3.4	2.7	4.68	5.36	4.63	4.85	5.70	7.10	-	-	-
Gender	1.8	-	2.4	5.03	-	6.20	6.04	9.12	7.03	6.31	4.57	7.19
Genotype relative risk (GRR)	5.1	4.3	4.7	8.56	6.75	9.11	6.53	12.54	5.64	5.57	6.26	5.34
Ileal/Colonic involvement*	n/a	n/a	n/a	5.50	-	5.87	4.94	6.40	5.81	n/a	n/a	n/a
Proportion of blood T-cells with memory phenotype (%)	2.0	-	2.4	-	-	-	8.71	5.77	9.24	5.19	-	6.87
Proportion of $CD4^+$ naïve T-cells expressing β7 integrin (%)	-	3.6	-	4.85	5.84	5.19	5.30	-	7.81	5.60	10.08	4.15
Undetermined	89.4	85.9	87.8	67.2	78.4	64.5	56.3	51.6	49.1	72.0	73.0	71.4

319 * Ileal/Colonic involvement in Crohn's patients used as a factor for the corresponding Sibling subjects. n/a denotes not applicable for between cohort or within
320 Healthy cohort analyses.

DISCUSSION

This is the first study to detail the mucosal microbiota of clinically and genetically well-characterised healthy siblings of CD patients, and to compare them with both their CD-affected siblings and healthy controls. Moreover, this study is unique in uncovering interactions of mucosal microbiota with genotype and features of the CD-risk phenotype. This manuscript is a significant advance on the preliminary account of the multidimensional risk phenotype previously described, which centred on qPCR sampling of faecal microbiota.[14] The current study not only focuses on the mucosal microbiota but also employs next-generation sequencing and advanced statistical analysis to reveal the complexity of the metacommunities in healthy siblings of CD patients. The core mucosal microbiota in siblings was characterised by lower diversity compared with controls, and lower abundance of *F. prausnitzii* made the greatest contribution to the dissimilarity between these two groups. Genetic CD-risk explained the highest proportion of microbial variance both between all three groups, and within the patient and sibling groups. These findings are unlikely to be confounded by cohabitation as only one patient cohabited with one sibling.

Although related healthy individuals are known to harbour similar gut microbiota,[19] the similarity in the microbiota between CD patients and their healthy siblings is of considerable pathogenic relevance. Previous studies have shown that when one sibling has CD, familial microbial similarity is disrupted, even in disease-discordant monozygotic twins.[30] Thus, microbial features which are similar between affected and unaffected siblings, but which are not present in low CD-risk healthy individuals, may be part of the CD-risk phenotype and therefore pertinent to CD pathogenesis. In order to discern these features associated with familial risk, comparison with healthy, unrelated individuals is essential.

The validity of the data presented is supported by the correlation between species-abundance and distribution, which is consonant with a coherent metacommunity structure and is similar to distributions described in other ecological communities.[15] This feature of community structure facilitated delineation of core species which are abundant and persistent, and allowed resolution of features of the mucosal microbiota without obfuscation from rare microbiota which may be highly variable, transient and scarce. A

significantly higher proportion of the microbiota in CD patients belonged to the rare group compared with healthy siblings and healthy controls. As described below this is at least in part attributable to loss of principal members of the core group, most notably Firmicutes.

Reduced microbial diversity is an almost universally reported feature of mucosal CD dysbiosis.[1] The current study reveals that core microbiota diversity is also lost in siblings of CD patients, indicating that this may be a fundamental step in CD pathogenesis. Reduced diversity may be an indicator of the health of human microbial communities, as it is reduced in a variety of disorders.[18-21] Lower diversity may be associated with incomplete occupation of ecological niches resulting in reduced resistance to pathogen colonisation; additionally a more restricted gut metagenome contains a lower array of genes which may result in the loss of key functions.

Lower diversity indicates altered mucosal microbial composition, and microbial composition in CD patients and healthy controls were significantly distinct from one another. In contrast, the composition of the whole and core microbiota in healthy siblings was not significantly different from either CD patients or healthy controls, indicating that from a microbial metacommunity perspective, siblings lie somewhere between patients and controls. The greater variability in the composition of the microbiota in at-risk siblings (illustrated by larger 95% concentration ellipse in Figure 3 (panel B)) probably reflects the range of CD-risk contained within this group, with siblings with higher CD-risk lying closer to or within the CD region. In addition, diversity was lower in core and rare microbiota in patients with ileocaecal resection/ right hemicolectomy, potentially explained by differences in disease phenotype, or the absence of the ileocaecal valve that would otherwise constitute a barrier between small and large intestinal microbiota.

Consonant with previous work highlighting the importance of *F. prausnitzii* in CD dysbiosis,[7, 12, 14] *F. prausnitzii* made the greatest contribution to the dissimilarity between CD patients and healthy control microbiota. The prominence of *F. prausnitzii* has biological significance as it is the only microbial factor shown to be predictive of the natural history of CD,[7] and response to treatment.[31] Strikingly, *F.*

prausnitzii was also the biggest contributor to the dissimilarity of the core mucosal microbiota between healthy siblings and healthy controls, establishing that mucosal *F. prausnitzii* not only correlates to the natural history of CD, but is also a key feature of the at-risk phenotype. Taken together these findings strongly support the hypothesis that depletion of *F. prausnitzii* is part of CD pathogenesis rather than consequent to established CD. Several mechanisms exist whereby *F. prausnitzii* and other Firmicutes may contribute to gut health, including the production of short-chain fatty acids (SCFAs),[32, 33] SCFA-independent, NFκB-mediated effects,[7] and via production of longer-chain fatty acids such as conjugated linoleic acid.[34]

The pathogenic role of reduced *F. prausnitzii* in CD has been questioned by a study describing increased mucosal *F. prausnitzii* in newly-diagnosed pediatric IBD.[35] However, whether increased abundance of *F. prausnitzii* is a distinctive feature of pediatric-onset IBD, with low *F. prausnitzii* being associated with later-onset CD, or whether the abundance of *F. prausnitzii* may bloom in childhood and then critically decline in those at risk of CD, may only be determined by longitudinal studies.

Other species contributing to the dissimilarity in the core mucosal microbiota between CD patients and healthy controls were congruent with species previously identified as characterising the CD dysbiosis, including a greater abundance of most Proteobacteria such as *E. fergusonii* and *Escherichia coli*. Similar species contributed to the dissimilarity between siblings and controls. However, the presence of *E. coli* was specific to CD mucosa, and therefore may be a feature of established CD rather than pathogenic. Features of the inflamed gut such as increased activity of nitric oxide synthases[36], or reduction in faecal butyrate producers which will result in a rise in pH, potentially favour the survival of organisms that are inhibited at acidic pH such as *E. coli*.[37]

GRR was the factor associated most strongly with the variation in the microbiota in both the between-group analysis, and analysis within each of the three groups. Although the proportion of variation in mucosal microbiota explained by GRR was small, it is nevertheless significant. The combination of loci used

394 to estimate GRR in the current study does not include more recently detected risk loci and can be expected
395 to account for a limited proportion of the genetic risk.[38] Therefore, these data will tend to underestimate
396 the effect of genotype. Furthermore, since other factors known to affect gut microbiota such as diet were
397 not controlled, this signal of the interaction between genotype and the mucosal microbiota is striking.

398 The direction of the vector in figure 3 illustrates that FC contributed to the axis separating patients from the
399 other two groups in the whole, core and rare microbiota, implying that microbial composition in CD is
400 partly associated with the degree of inflammation. This would support the hypothesis that CD-specific
401 elements of the dysbiosis may be consequent to intestinal inflammation, through mechanisms such as the
402 enhanced survival of *E. coli* in an inflamed environment as proposed above.

403 When each group was considered separately, the effect of each factor in different groups could be
404 compared. Several factors were significant in all groups (GRR, gender, proportion of $CD4^+$ naïve T-cells
405 expressing β7 integrin). Other factors were significant in patients and siblings but not controls: FC and
406 blood naïve $CD4^+$ T-cell concentration were significant only in patients and siblings, whereas age was
407 significant only in controls. Disease phenotype was significant in explaining microbial variation within the
408 CD group as would be predicted from previous studies.[30] However; we have also demonstrated that for
409 healthy siblings, disease site in their affected relative was significantly associated with the variation in their
410 own microbiota. This would suggest that specific risk phenotypes are associated with different disease
411 phenotypes.

412 Overall these factors accounted for a higher proportion of the variance in the microbial composition in
413 siblings, compared with controls or patients, indicating that this multidimensional risk phenotype is specific,
414 and that in low CD-risk individuals the microbial composition is associated with other factors, such as age.
415 Furthermore, it would appear that in CD the influence of factors associated with the original risk phenotype
416 is obfuscated by established CD and its surgical and medical management.

417 **CONCLUSION**

Healthy siblings of CD patients, who themselves have elevated risk of CD, have a dysbiosis of the core mucosal microbiota characterised by reduced diversity and loss of Firmicutes, notably *F. prausnitzii*. Genotype determines a proportion of the at-risk mucosal microbial phenotype. Notwithstanding the limited extent to which known loci account the observed CD-risk,[39] it is also clear that the sibling risk goes beyond genotype and that non-genetic factors within families contribute to the development of an at-risk microbiota. How and why patients and their siblings acquire the microbiota that marks out this risk is not known. However, knowledge of the at-risk microbial phenotype illuminates possible pathways in CD pathogenesis and raises the prospect of intervention to impact human health and influence disease risk.

Acknowledgements

We gratefully acknowledge and thank the charity Core who funded a research fellowship for CRH which made this study possible.

CvdG and LC were supported by the UK Natural Environment Research Council (grant NE/H019456/1)

FIGURE LEGENDS

Fig. 1. The distribution and abundance of bacterial species within microbiota samples within the (a) CD, (b) siblings, and (c) healthy control cohort metacommunities. Given is the number of mucosal samples for which each bacterial taxon was observed to occupy, plotted against the mean abundance across all samples ((a) n =21, r2 =0.62, F1, 227 =366.9, P < 0.0001; (b) n =17, r2 =0.71, F1, 259 =590.1, P < 0.0001; and (c) n =19, r2 =0.68, F1, 258 =552.6, P < 0.0001). Core species were defined as those that fell within the upper quartile (dashed lines), and rare species defined as those that did not.

Fig. 2. Diversity of whole, core and rare microbiota within the CD (black columns), siblings (grey), and healthy (white)-control cohorts. Given are three indices of diversity; Species richness (*S**), Simpson's index of diversity (1-*D*), and Shannon-Wiener index of diversity *(H')*. Error bars represent the standard deviation of the mean (CD *n* =21, siblings *n* =17, and healthy *n* =19). Asterisks denote significant differences in comparisons of diversity at the P < 0.05 level determined by two sample *t*-tests.

Fig. 3. Analysis of similarities (ANOSIM) of whole, common, and rare microbiota between subject cohorts. Given is the ANOSIM test statistic (*R*) and probability (*P*) that two compared groups are significantly different at the *P* < 0.05 level (* denotes *P* < 0.001 and ** *P* < 0.0001). ANOSIM *R* and *P* values were generated using the Bray-Curtis measure of similarity. *R* scales from +1 to -1. +1 indicates that all the most similar samples are within the same groups. *R* = 0 occurs if the high and low similarities are perfectly mixed and bear no relationship to the group. A value of -1 indicates that the most similar samples are all outside of the groups.

Fig. 4. Canonical correspondence biplots for (a) whole, (b) core, and (c) rare microbiota. Red crosses represent microbiota samples from the CD cohort, yellow filled triangles for the siblings cohort, and green diamonds for the healthy cohort. In each instance, the 95 % concentration ellipses are given for the CD (red), siblings (yellow), and healthy (green) cohort microbiota. Biplot lines for clinical variables that significantly accounted for variation within the microbiota at the $P < 0.05$ level (see Table 3) show the direction of increase for each variable, and the length of each line indicates the degree of correlation with the ordination axes. CCA field labels: Calprotectin, Gender, "T-cells" – Proportion of blood T-cell with memory phenotype (%), "$CD4^+$ T-cells" – Blood concentration of naïve $CD4^+$ T-cells (cells /ml), "β7 integrin" – Proportion of CD4 naïve T-cells expressing β7 integrin (%),"GRR" – genotype relative risk, (cumulative genotype relative risk (GRR) for each participant was calculated across 72 CD-risk loci (detected using the Illumina Infinium Immunochip), participants were categorised into reduced, average, elevated or high genotype risk with reference to a population distribution model of CD-risk). Percentage of community variation explained by each axis is given in parentheses.

REFERENCE LIST

1. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 2014;146:1489-1499.

2. Franke A, McGovern DP, Barrett JC *et al*. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118-25

3. Anderson CA, Boucher G, Lees CW *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet. 2011;43:246-52.

4. Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491:119-24.

5. Sellon RK, Tonkonogy S, Schultz M *et al*. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infect Immun* 1998;66:5224-5231.

6. Harper PH, Lee EC, Kettlewell MG *et al*. Role of the faecal stream in the maintenance of Crohn's colitis. *Gut* 1985;26:279-284.

7. Sokol H, Pigneur B, Watterlot L *et al*. *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 2008;105:16731-16736.

8. Ghouri YA, Richards DM, Rahimi EF *et al.* Systematic review of randomized controlled trials of probiotics, prebiotics, and synbiotics in inflammatory bowel disease. Clin Exp Gastroenterol. 2014;9:473-87.

9. Benjamin JL, Hedin CR, Koutsoumpas A *et al*. Randomised, double-blind, placebo-controlled trial of fructo-oligosaccharides in active Crohn's disease. *Gut* 2011;60:923-929.

10. Calkins BM, Mendeloff AI. Epidemiology of inflammatory bowel disease. *Epidemiol Rev* 1986;8:60-91.

11. Halfvarson J. Genetics in twins with Crohn's disease: less pronounced than previously believed? *Inflamm Bowel Dis* 2011;17:6-12.

12. Hedin CR, Stagg AJ, Whelan K *et al*. Family studies in Crohn's disease: new horizons in understanding disease pathogenesis, risk and prevention. *Gut* 2012;61:311-318.

13. Joossens M, Huys G, Cnockaert M *et al*. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 2011;60:631-637.

14. Hedin CR, McCarthy NE, Louis P *et al*. Altered intestinal microbiota and blood T cell phenotype are shared by patients with Crohn's disease and their unaffected siblings. *Gut* 2014;63:1578-1586.

15. van der Gast CJ, Walker AW, Stressmann FA *et al*. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* 2011;5:780-791.

16. Sogin ML, Morrison HG, Huber JA *et al*. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* 2006;103:12115-12120.

17. Li K, Bihan M, Yooseph S *et al*. Analyses of the microbial diversity across the human microbiome. *PLoS One* 2012;7:e32118.

18. Le CE, Nielsen T, Qin J *et al*. Richness of human gut microbiome correlates with metabolic markers. *Nature* 2013;500:541-546.

19. Turnbaugh PJ, Hamady M, Yatsunenko T *et al*. A core gut microbiome in obese and lean twins. *Nature* 2009;457:480-484.

20. Ahn J, Sinha R, Pei Z *et al*. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst* 2013;105:1907-1911.

21. Ismail IH, Oppedisano F, Joseph SJ *et al*. Reduced gut microbial diversity in early life is associated with later development of eczema but not atopy in high-risk infants. *Pediatr Allergy Immunol* 2012;23:674-681.

511 22. Biedermann L, Zeitz J, Mwinyi J *et al*. Smoking cessation induces profound changes in the composition
512 of the intestinal microbiota in humans. *PLoS One* 2013;8:e59260.

513 23. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. *Arthritis Res Ther* 2011;13:101.

514 24. Crouch DJ, Goddard GH, Lewis CM. REGENT: a risk assessment and classification algorithm for genetic
515 and environmental factors. *Eur J Hum Genet* 2013;21:109-111.

516 25. Goddard GH, Lewis CM. Risk categorization for complex disorders according to genotype relative risk
517 and precision in parameter estimates. *Genet Epidemiol* 2010;34:624-632.

518 26. Rogers GB, Cuthbertson L, Hoffman LR *et al*. Reducing bias in bacterial community analysis of lower
519 respiratory infections. *ISME J* 2013;7:697-706.

520 27. Rogers GB, Kozlowska J, Keeble J *et al*. Functional divergence in gastrointestinal microbiota in physically-
521 separated genetically identical mice. *Sci Rep* 2014;4:5437.

522 28. Dowd SE, Wolcott RD, Sun Y, *et al.* Polymicrobial nature of chronic diabetic foot ulcer biofilm infections
523 determined using bacterial tag encoded FLX amplicon pyrosequencing (bTEFAP). *PLoS One* 2008;3:e3326.

524 29. R Core Team.R Foundation for Statistical Computing. Vienna A. R: A Language and Environment for
525 Statistical Computing. 2014.

526 30. Willing B, Dicksved J, Halfvarson J *et al*. A pyrosequencing study in twins shows that GI microbial profiles
527 vary with inflammatory bowel disease phenotypes. *Gastroenterology* 2010;39:1844-1854.

528 31. Rajca S, Grondin V, Louis E *et al.* Alterations in the intestinal microbiome (dysbiosis) as a predictor of
529 relapse after infliximab withdrawal in Crohn's disease. *Inflamm Bowel Dis*. 2014;20:978-86.

530 32. Macfarlane GT, Macfarlane S. Fermentation in the human large intestine: its physiologic consequences
531 and the potential contribution of prebiotics. *J Clin Gastroenterol* 2011;45 Suppl:S120-S127.

532 33. Vinolo MA, Rodrigues HG, Nachbar RT *et al*. Regulation of inflammation by short chain fatty acids.
533 *Nutrients* 2011;3:858-876.

534 34. Bassaganya-Riera J, Hontecillas R, Horne WT *et al*. Conjugated linoleic acid modulates immune
535 responses in patients with mild to moderately active Crohn's disease. *Clin Nutr* 2012.

536 35. Hansen R, Russell RK, Reiff C *et al*. Microbiota of de-novo pediatric IBD: increased Faecalibacterium
537 prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am J Gastroenterol*
538 2012;107:1913-1922.

539 36. Rachmilewitz D, Karmeli F, Okon E *et al*. Experimental colitis is ameliorated by inhibition of nitric oxide
540 synthase activity. *Gut* 1995;37:247-255.

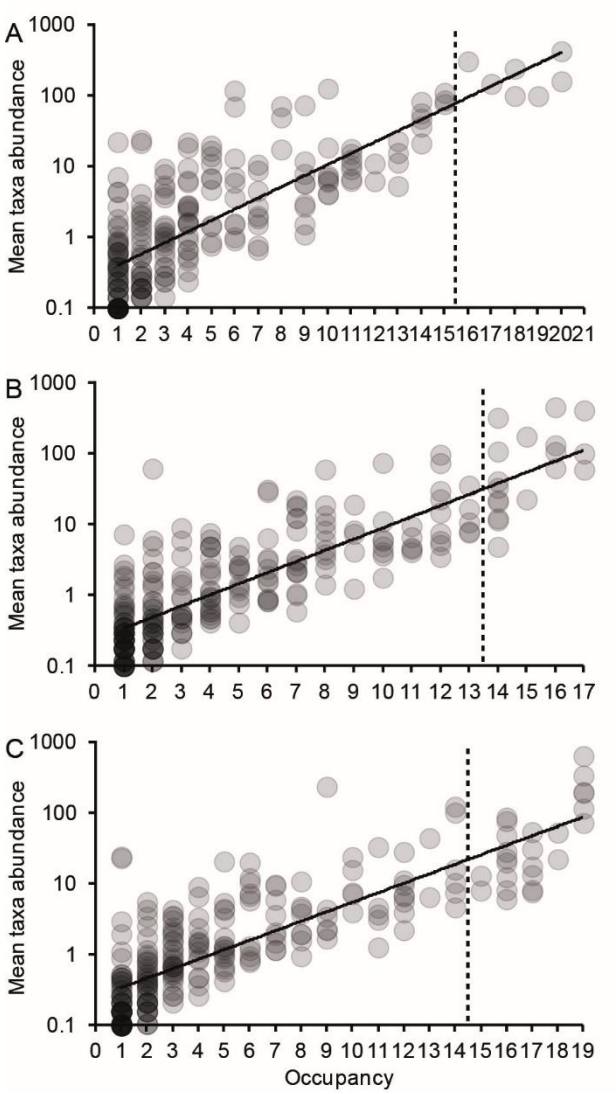
541 37. Duncan SH, Louis P, Thomson JM *et al*. The role of pH in determining the species composition of the
542 human colonic microbiota. *Environ Microbiol* 2009;11:2112-2122.

543 38. Park JH, Wacholder S, Gail MH *et al*. Estimation of effect size distribution from genome-wide association
544 studies and implications for future discoveries. *Nat Genet* 2010;42:570-575.

545 39. Manolio TA, Collins FS, Cox NJ *et al*. Finding the missing heritability of complex diseases. *Nature*
546 2009;461:747-753.

547

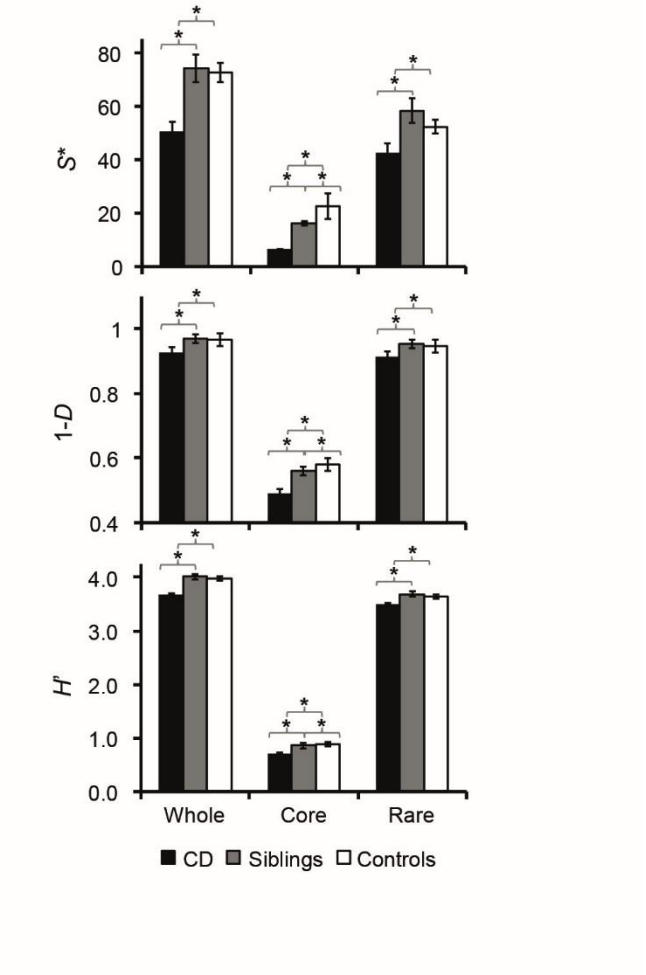
548 Figure 1



549

550

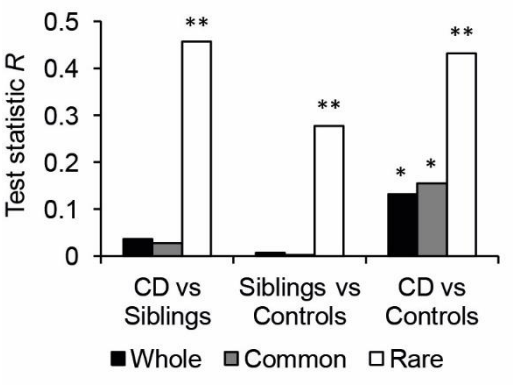
551 Figure 2



552

553

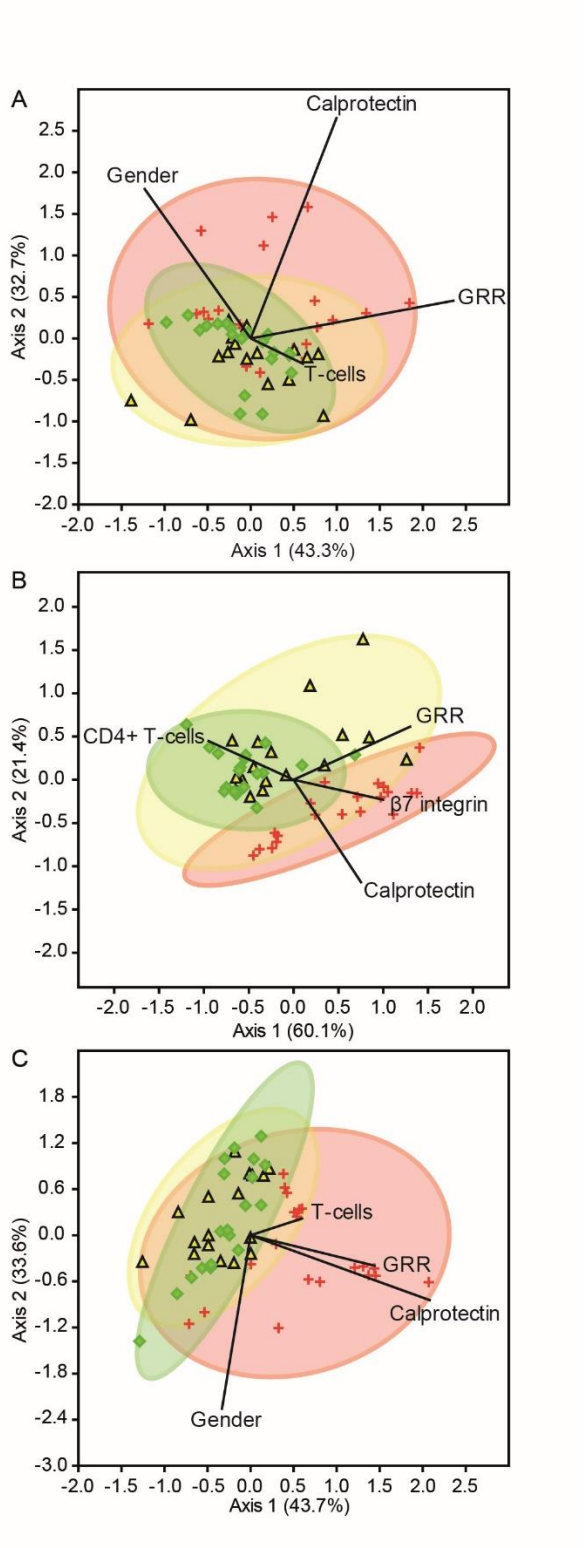
554 Figure 3



555

556

557 Figure 4



558

559

560 **Supplementary table S1.** List of exclusion criteria by group

All participants	Patients	Siblings	Controls
Unable to consent due to mental illness/ dementia/ learning disability	Evidence of active CD as defined by a CDAI of greater than 150	Previous diagnosis of IBD	Previous diagnosis of IBD
Current infection with an enteric pathogen	Purely perianal CD	Symptoms fulfilling Rome III criteria for IBS	Symptoms fulfilling Rome III criteria for IBS
Use of antibiotics within the last month	Change in dose of oral steroids within the last 4 weeks		A first or second degree relative with IBD
Consumption of any probiotic or prebiotic within the last month	Dose of steroids exceeding 10mg prednisolone per day or equivalent		
Pregnancy or lactation	Change in dose of oral 5-ASA products within the last 4 weeks		
Participant requiring hospitalization	Commencement of azathioprine or methotrexate within the last 4 months, or change in dose of these drugs within the last 4 weeks		
Significant hepatic, renal, endocrine, respiratory, neurological or cardiovascular disease as determined by the principal investigator	Infusion of biological therapies (e.g. infliximab) within the last 3 months*		
A history of cancer with a disease free state of less than two years	Use of rectal 5-ASA or steroids within the last 2 weeks		
CRP greater than 5mg/L at screening, as measured by the local laboratory	Use of NSAIDs within the last 2 weeks		
	Imminent need for surgery		
	Short bowel syndrome		
	Previous proctocolectomy		

561 * No patient had been previously exposed to biological therapies

562

Supplementary table S2 Bacterial species identified from biopsy samples collected from subject within the Crohn's (*n* = 21), Siblings (*n* = 17), and Healthy (*n* = 19) cohorts. Species-level identities of detected taxa are reported here. However, given the length of the ribosomal sequences analysed, these identities should be considered putative. C and R denote core or rare species group membership within each cohort (highlighted in green and yellow, respectively).

Phylum	Class	Family	Taxon Name	Crohn's	Siblings	Healthy
Actinobacteria	Actinobacteria	Actinomycetaceae	*Actinomyces naeslundii*			R
			Varibaculum cambriense		R	R
		Corynebacteriaceae	*Corynebacterium afermentans*		R	
			Corynebacterium amycolatum	R	R	R
			Corynebacterium aurimucosum		R	
			Corynebacterium durum	R		
			Corynebacterium glucuronolyticum			R
			Corynebacterium imitans		R	
			Corynebacterium jeikeium		R	R
			Corynebacterium mucifaciens		R	
			C. pseudogenitalium			R
		Geodermatophilaceae	*Blastococcus saxobsidens*		R	
		Intrasporangiaceae	*Janibacter*		R	
		Microbacteriaceae	*Agrococcus jejuensis*	R		
			Arthrobacter agilis		R	
			Microbacterium barkeri		R	R
			Microbacterium paraoxydans		R	R
			Zimmermannella bifida		R	
		Micrococcaceae	*Micrococcus luteus*	R	R	R
		Mycobacteriaceae	*Mycobacterium llatzerense*		R	
		Nocardiaceae	*Rhodococcus globerulus*	R	R	
		Propionibacteriaceae	*Microlunatus aurantiacus*			R
			Propionibacterium acnes	R	R	R
			P. granulosum		R	
		Pseudonocardiaceae	*Saccharopolyspora hirsuta*	R	R	R
		Streptomycetaceae	*Streptomyces thermovulgaris*	R		
		Williamsiaceae	*Williamsia muralis*			R
		Bifidobacteriaceae	*Bifidobacterium adolescentis*	R	R	R
			Bifidobacterium bifidum	R	R	R
			Bifidobacterium breve	R		R
			Bifidobacterium longum	R	R	R
			B. pseudocatenulatum	R	R	R
			Bifidobacterium saeculare	R		
			Gardnerella vaginalis			R
		Coriobacteriaceae	*Adlercreutzia equolifaciens*	R	R	R
			Atopobium vaginae		R	R
			Collinsella aerofaciens	R	R	C
			Collinsella stercoris		R	R
			Eggerthella hongkongensis	R		R
			Eggerthella lenta	R	R	R
			Eggerthella sinensis			R
			Olsenella			R
			Slackia	R		R
Bacteroidetes	Bacteroidia	Bacteroidaceae	*Bacteroides acidifaciens*		R	R
			Bacteroides barnesiae			R
			Bacteroides caccae	R	R	R

			Bacteroides capillosus	R	R	R
			Bacteroides cellulosilyticus	R	R	R
			Bacteroides coprocola	R	R	

Table S1 continued

Phylum	Class	Family	Taxon Name	Crohn's	Siblings	Healthy
			Bacteroides coprophilus		R	R
			Bacteroides dorei	R	C	R
			Bacteroides eggerthii	R	R	R
			Bacteroides faecis	R	R	R
			Bacteroides finegoldii		R	R
			Bacteroides fragilis	R	R	R
			Bacteroides intestinalis	R	R	R
			Bacteroides massiliensis	R	R	R
			Bacteroides nordii			R
			Bacteroides ovatus	R	R	R
			Bacteroides plebeius	R	R	R
			Bacteroides salyersiae	R	R	R
			Bacteroides stercoris	R	R	R
			Bacteroides thetaiotaomicron	R	R	R
			Bacteroides uniformis	R	C	C
			Bacteroides vulgatus	C	C	C
			Bacteroides xylanisolvens	R	R	R
		Porphyromonadaceae	*Dysgonomonas gadei*	R	R	C
			Parabacteroides distasonis	R	C	C
			Parabacteroides goldsteinii	R	R	R
			Parabacteroides johnsonii	R	R	R
			Parabacteroides merdae	R	R	C
			Porphyromonas levii	R	R	R
			Porphyromonas somerae	R	R	R
			Porphyromonas uenonis		R	R
			Tannerella forsythia	R	R	R
		Prevotellaceae	*Prevotella baroniae*	R	R	R
			Prevotella bergensis		R	R
			Prevotella bivia	R	R	R
			Prevotella buccae		R	
			Prevotella buccalis	R	R	R
			Prevotella copri	R	R	R
			Prevotella corporis	R	R	R
			Prevotella dentasini			R
			Prevotella disiens	R	R	R
			Prevotella histicola		R	R
			Prevotella loescheii			R
			Prevotella melaninogenica		R	R
			Prevotella oulorum		R	
			Prevotella pallens			R
			Prevotella pleuritidis		R	
			Prevotella ruminicola		R	R
			Prevotella stercorea	R	R	R
		Rikenellaceae	*Alistipes finegoldii*	R	R	R
			Alistipes massiliensis	R	R	R
			Alistipes putredinis	R	R	R
	Cytophagia	Cytophagaceae	*Hymenobacter*		R	R
		Flammeovirgaceae	*Rapidithrix thailandica*		R	

	Flavobacteriia	Flavobacteriaceae	*Chryseobacterium hominis*	R	R	
			Chryseobacterium joostei	R	R	
	Sphingobacteriia	Chitinophagaceae	*Chitinophaga arvensicola*	R		R
Deinococcus-Thermus	Deinococci	Deinococcaceae	*Deinococcus proteolyticus*	R		
Firmicutes	Bacilli	Alicyclobacillaceae	*Alicyclobacillus acidoterrestris*	R		
			Alicyclobacillus vulcanalis			R

Table S1 continued

Phylum	Class	Family	Taxon Name	Crohn's	Siblings	Healthy
		Bacillaceae	*Anoxybacillus kestanbolensis*		R	R
			Bacillus cereus	R	C	R
			Bacillus fordii		R	
			Geobacillus stearothermophilus		R	
		Paenibacillaceae	*Paenibacillus humicus*		R	
		Planococcaceae	*Viridibacillus arvi*		R	
		Staphylococcaceae	*Jeotgalicoccus halotolerans*		R	R
			Macrococcus caseolyticus	R		
			Salinicoccus roseus		R	
			Staphylococcus aureus	R		R
			Staphylococcus epidermidis	R	R	R
			Staphylococcus hominis	R		R
		Aerococcaceae	*Abiotrophia defectiva*	R	R	R
			Aerococcus viridans			R
			Facklamia ignava		R	
			Facklamia languida		R	
		Carnobacteriaceae	*Marinilactibacillus psychrotolerans*	R		R
		Enterococcaceae	*Enterococcus avium*		R	
			Enterococcus cecorum	R	R	R
			Enterococcus durans	R		
			Enterococcus faecium	R		R
			Enterococcus lactis	R		R
		Lactobacillaceae	*Lactobacillus alimentarius*		R	
			Lactobacillus crispatus	R		R
			Lactobacillus delbrueckii		R	
			Lactobacillus gasseri	R	R	R
			Lactobacillus iners	R		R
			Lactobacillus intestinalis	R	R	R
			Lactobacillus manihotivorans	R		
			Lactobacillus mucosae	R	R	R
			Lactobacillus reuteri		R	
			Lactobacillus vaccinostercus	R		
			Lactobacillus zeae	R		
			Pediococcus acidilactici	R		
		Leuconostocaceae	*Weissella cibaria*	R	R	
			Weissella confusa	R		
		Streptococcaceae	*Lactococcus lactis*		R	R
			Streptococcus anginosus	R	R	R
			Streptococcus bovis	R		
			Streptococcus dysgalactiae			R
			Streptococcus gallinaceus			R
			Streptococcus gordonii			R
			Streptococcus mitis	R	R	R
			Streptococcus oralis	R	R	R
			Streptococcus parasanguinis	R	R	R

			Streptococcus pluranimalium			R
			Streptococcus thermophilus	R	R	R
	Clostridia	Clostridiaceae	*Alkaliphilus transvaalensis*	R	R	R
			Clostridium amygdalinum	R	R	R
			Clostridium asparagiforme	R	R	R
			Clostridium bartlettii	R	R	R
			Clostridium bolteae	R	C	C
			Clostridium butyricum	R		
			Clostridium celerecrescens	R	R	R

Table S1 continued

Phylum	Class	Family	Taxon Name	Crohn's	Siblings	Healthy
			Clostridium chauvoei		R	R
			Clostridium citroniae	R	R	R
			Clostridium cocleatum	R	R	R
			Clostridium difficile	R		
			Clostridium disporicum	R		
			Clostridium ghonii	R		R
			Clostridium hathewayi	R	R	R
			Clostridium hylemonae			R
			Clostridium indolis	R	R	C
			Clostridium innocuum	R	R	R
			Clostridium intestinale			R
			Clostridium lactatifermentans	R	R	R
			Clostridium lavalense	R	R	R
			Clostridium leptum	R	R	R
			C. methoxybenzovorans	R	R	R
			Clostridium methylpentosum	R	R	C
			Clostridium orbiscindens	R	C	C
			Clostridium paraputrificum	R		
			Clostridium perfringens	R		
			Clostridium puniceum	R	R	R
			Clostridium ramosum	R	R	R
			Clostridium saccharolyticum	R	R	R
			Clostridium scindens	R		
			Clostridium spiroforme	R	R	R
			Clostridium stercorarium	R		
			Clostridium straminisolvens	R	R	R
			Clostridium symbiosum	R	C	R
			Clostridium termitidis		R	R
			Clostridium thermocellum	R	R	R
			Clostridium xylanolyticum	R	R	R
		Clostridiales Family XI	*Anaerococcus hydrogenalis*	R	R	R
			Anaerococcus lactolyticus			R
			Anaerococcus octavius	R	R	R
			Anaerococcus prevotii	R	R	R
			Finegoldia magna	R	R	R
			Parvimonas micra	R	R	R
			Peptoniphilus asaccharolyticus	R	R	R
			Peptoniphilus ivorii	R	R	R
			Tissierella praeacuta	R	R	R
		Clostridiales Family XIII	*Eubacterium infirmum*	R	R	R
			Eubacterium minutum		R	
			Eubacterium sulci	R	R	R

			Mogibacterium pumilum	R	R	R
		Clostridiales Family XVIII	*Symbiobacterium thermophilum*		R	
		Eubacteriaceae	*Eubacterium rectale*	C	C	C
			Eubacterium siraeum	R	R	R
		Lachnospiraceae	*Anaerostipes caccae*		R	
			Blautia producta	R	C	C
			Coprococcus catus	R	R	R
			Coprococcus comes	R	R	R
			Coprococcus eutactus	R	R	C
			Dorea formicigenerans	R	R	R
			Lachnospira pectinoschiza	R	R	R
			Moryella indoligenes		R	R

Table S1 continued

Phylum	Class	Family	Taxon Name	Crohn's	Siblings	Healthy
			Pseudobutyrivibrio fibrisolvens	R	R	R
			Pseudobutyrivibrio ruminis	R	R	R
			Roseburia faecis	R	C	C
			Roseburia hominis	R	R	R
			Roseburia intestinalis	R	R	R
		Peptococcaceae	*Desulfitobacterium hafniense*		R	R
		Peptostreptococcaceae	*Eubacterium yurii*		R	R
			Peptostreptococcus anaerobius	R	R	R
			Peptostreptococcus stomatis		R	
		Ruminococcaceae	*Anaerotruncus colihominis*	R	R	R
			Faecalibacterium prausnitzii	C	C	C
			Oscillospira guilliermondii	R	C	C
			Ruminococcus albus	R	R	R
			Ruminococcus bromii	R	C	C
			Ruminococcus callidus	R	R	R
			Ruminococcus flavefaciens	R	R	R
			Ruminococcus gnavus	C	C	C
			Ruminococcus obeum	R	R	C
			Ruminococcus torques	R	R	C
		Syntrophomonadaceae	*Syntrophomonas curvata*			R
	Erysipelotrichi	Erysipelotrichaceae	*Catenibacterium mitsuokai*	R	R	R
			Coprobacillus cateniformis	R	R	R
			Erysipelothrix rhusiopathiae	R		
			Eubacterium biforme	R	R	R
			Eubacterium cylindroides		R	R
			Holdemania filiformis	R	R	R
		Erythrobacteraceae	*Solobacterium moorei*		R	R
	Negativicutes	Acidaminococcaceae	*Acidaminococcus fermentans*	R	R	R
			Acidaminococcus intestini	R	R	R
			Phascolarctobacterium	R	R	R
		Veillonellaceae	*Dialister invisus*	R	R	C
			Dialister micraerophilus	R	R	R
			Dialister pneumosintes	R	R	
			Dialister succinatiphilus	R	R	R
			Megamonas hypermegale		R	R
			Megasphaera elsdenii	R	R	R
			Mitsuokella jalaludinii		R	R
			Mitsuokella multacida	R		R
			Sporomusa aerivorans	R		

			Veillonella dispar	R	R	R
			Veillonella parvula	R	R	R
			Victivallis vadensis		R	R
Fusobacteria	Fusobacteria	Fusobacteriaceae	*Fusobacterium gonidiaformans*		R	R
			Fusobacterium nucleatum		R	
			Fusobacterium periodonticum	R	R	R
			Fusobacterium ulcerans	R	R	R
		Leptotrichiaceae	*Leptotrichia goodfellowii*	R		
			Sneathia amnionii	R	R	
			Sneathia sanguinegens			R
Planctomycetes	Planctomycetia	Planctomycetaceae	*Gemmata obscuriglobus*	R		
Proteobacteria	Alphaproteobacteria	Caulobacteraceae	*Asticcacaulis excentricus*		R	
			Brevundimonas nasdae			R
		Bradyrhizobiaceae	*Nitrobacter vulgaris*			R
			Rhodopseudomonas	R	R	R

Table S1 continued

Phylum	Class	Family	Taxon Name	Crohn's	Siblings	Healthy
		Methylobacteriaceae	*Methylobacterium tardum*	R	R	
		Methylocystaceae	*Methylocystis parvus*			R
		Phyllobacteriaceae	*Mesorhizobium loti*			R
			Phyllobacterium myrsinacearum	R	R	R
		Rhodobacteraceae	*Labrenzia aggregata*		R	
			Loktanella marincola		R	
			Paracoccus alcaliphilus	R	R	
			Paracoccus koreensis	R	R	
			Paracoccus marcusii	R		R
			Paracoccus yeei	R	R	R
			Rubellimicrobium	R	R	
		Rhodospirillaceae	*Rhodocista pekingensis*			R
		Anaplasmataceae	*Neorickettsia risticii*	R	R	R
		Erythrobacteraceae	*Porphyrobacter tepidarius*		R	
		Sphingomonadaceae	*Sphingobium yanoikuyae*			R
			Sphingomonas asaccharolytica		R	R
			Sphingomonas faeni			R
	Betaproteobacteria	Comamonadaceae	*Acidovorax facilis*	R	R	R
		Oxalobacteraceae	*Massilia timonae*	R	R	R
		Sutterellaceae	*Sutterella parvirubra*	R	R	R
			Sutterella wadsworthensis	R	C	C
		Burkholderiaceae	*Pandoraea pulmonicola*			R
			Ralstonia pickettii	R	R	R
		Comamonadaceae	*Hydrogenophaga taeniospiralis*		R	
			Variovorax paradoxus	R		
		Hydrogenophilaceae	*Hydrogenophilus*			R
		Neisseriaceae	*Microvirgula aerodenitrificans*		R	
			Neisseria mucosa			R
			Neisseria subflava			R
		Burkholderiaceae	*Cupriavidus taiwanensis*	R	R	
		Comamonadaceae	*Comamonas denitrificans*	R	R	
			Comamonas terrigena		R	
			Comamonas testosteroni	R	R	
		Rhodocyclaceae	*Dechloromonas*		R	
	Deltaproteobacteria	Desulfovibrionaceae	*Bilophila wadsworthia*		R	R
			Desulfovibrio aespoeensis	R		

			Desulfovibrio piger	R	R	R
		Myxococcaceae	*Anaeromyxobacter dehalogenans*	R		
	Epsilonproteobacteria	Campylobacteraceae	*Campylobacter concisus*	R		
			Campylobacter faecalis			R
			Campylobacter hominis	R	R	R
			Campylobacter mucosalis		R	
			Campylobacter ureolyticus	R	R	R
	Gammaproteobacteria	Aeromonadaceae	*Aeromonas media*	R	R	R
		Succinivibrionaceae	*Succinivibrio*			R
		Shewanellaceae	*Shewanella putrefaciens*		R	R
		Enterobacteriaceae	*Shigella boydii*	R	R	R
			Shigella dysenteriae	R	R	R
			Shigella flexneri	C	C	C
			Shigella sonnei	R	R	R
			Citrobacter freundii	R	R	R
			Cronobacter muytjensii	R	R	R
			Cronobacter turicensis	R		
			Enterobacter hormaechei	R	R	R

Table S1 continued

Phylum	Class	Family	Taxon Name	Crohn's	Siblings	Healthy
			Escherichia coli	C	R	R
			Escherichia fergusonii	C	C	C
			Klebsiella granulomatis	R	R	R
			Klebsiella pneumoniae	R	R	R
			Klebsiella variicola	R	R	R
			Kluyvera ascorbata	R		
			Morganella morganii	R		
			Raoultella planticola	R		R
		Oceanospirillaceae	*Marinospirillum insulare*		R	
		Pasteurellaceae	*Actinobacillus porcitonsillarum*			R
			Haemophilus parainfluenzae	R	R	R
		Moraxellaceae	*Acinetobacter johnsonii*	R	R	R
			Acinetobacter junii	R	R	R
			Acinetobacter lwoffii	R		R
			Acinetobacter radioresistens	R		
			Acinetobacter schindleri		R	
			Enhydrobacter aerosaccus		R	R
		Pseudomonadaceae	*Pseudomonas aeruginosa*	R	R	R
			Pseudomonas geniculata	R	R	C
			Pseudomonas mendocina		R	
			Pseudomonas putida	R	R	R
			Pseudomonas veronii		R	R
		Xanthomonadaceae	*Stenotrophomonas*		R	R
			Thermomonas dokdonensis			R
			Xanthomonas vesicatoria	R	R	R
Spirochaetes	Spirochaetes	Brachyspiraceae	*Brachyspira aalborgi*			R
Synergistetes	Synergistia	Synergistaceae	*Jonquetella anthropi*		R	
			Pyramidobacter piscolens		R	R
			Synergistes jonesii			R
Tenericutes	Mollicutes	Acholeplasmataceae	*Acholeplasma*		R	
			Acholeplasma parvum		R	R
Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiaceae	*Akkermansia muciniphila*	R	R	R

Supplementary methods

Peripheral blood T-cell flow cytometry

The fluorescently labeled antibodies used were: anti-CD3 Pacific Blue (clone OKT3, Biolegend, San Diego, CA,USA), anti-CD45RA PE-Cy7 (clone L48, BD Bioscience, Franklin Lakes, NJ, USA), anti-CD8 PerCP-Cy5.5 (clone SK1, BD Bioscience) and anti-CD4 APC (clone RPA-T4, BD Bioscience), anti-β7 PE (clone FIB504, BD Pharmingen).

Isotype-matched controls for mIgG1κ PE-Cy7 (clone MOPC-21, BD Pharmingen), rat IgG2a PE (clone R35-95, BD Bioscience), mIgG1 PE (clone MOPC-21, BD Bioscience), rIgM FITC (clone R4-22, BD Pharmingen) and mIgG1 FITC (clone MOPC-21, BD Pharmingen) were used to set positive and negative regions for gating during analysis. Anti-CD8 FITC (clone LT8, AbD Serotec, Kidlington, UK), anti-β7 PE (clone FIB504, BD Pharmingen), anti-CD8 PerCP-Cy5.5 (clone SK1, BD Bioscience), anti-CD45RA PE-Cy7 (clone L48, BD Bioscience) anti-CD3 PB (clone OKT3, Biologend) and anti-CD4 APC (clone RPA-T4, BD Bioscience) conjugated antibodies were used for off-line compensation.

Gut mucosal microbiota

DNA extraction protocol

Biopsy DNA extraction was carried out using a phenol/chloroform based method, as follows: Guanidinium thiocyanate–EDTA–sarkosyl (500 µL) and PBS (500 µL), pH 8.0, were added to biopsy samples. Cell disruption was achieved using a Fastprep-24 Instrument (MP Biomedicals Europe, Illkirch, France) 6.5 m/s, 60 s, followed by incubation at 90 °C for 1 min and −20 °C for 5 min. Cell debris was pelleted by centrifugation at 12 000 × g for 2 min at 4 °C. Supernatant was transferred to a fresh microfuge tube. NaCl (to a final concentration of 0.5 mol/L and polyethylene glycol (to a final concentration of 15%) were added and DNA precipitated at 4 °C for 30 min. DNA was pelleted by centrifugation at 12 000 × g for 2min at 4 °C and resuspended in 300 µL of sterile distilled water. Samples were heated at 90 °C for 30 s and vortexed.

Phenol/chloroform (1:1) (300 µL) was added, and samples were vortexed for 20 s before centrifugation at 12 000 × g at 4 °C for 3min. The upper phase was then transferred to a fresh microfuge tube. Total DNA was then precipitated by the addition of an equal volume of isopropanol, a 0.1-volume 10 mol/L ammonium acetate, and 1 µL of GenElute linear polyacrylamide (Sigma-Aldrich, Gillingham, UK) and incubated at −20 °C for 25 min. DNA was pelleted by centrifugation at 12 000 × g at 4 °C for 5 min. Pelleted DNA was then washed 3 times in 70% ethanol, dried, and resuspended in 50 µL of sterile distilled water. DNA extracts were quantified using the Picodrop Microlitre Spectrophotometer (GRI, Braintree, UK). Negative controls, consisting of sterile water, were included in the PMA treatment, DNA extraction, and PCR amplification steps.

16S rRNA gene sequencing

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed as described previously using Gray28F 5'-TTTGATCNTGGCTCAG-3' and Gray519r 5'-GTNTTACNGCGGCKGCTG-3').[1] A single-step 30 cycle PCR using HotStarTaq Plus Master Mix Kit (Qiagen, Valencia, CA) performed under the following conditions: 94°C for 5minutes, followed by 28 cycles of: 94°C for 30 seconds, 53°C for 40 seconds, and 72°C for 1 minute. Amplification was followed by a final elongation step at 72°C for 5 minutes. Following PCR, all amplicon products from different samples were mixed in equal concentrations and purified using Agencourt Ampure beads (Agencourt Bioscience Corporation, MA, USA). Samples were sequenced utilizing Roche 454 FLX titanium instruments and reagents following manufacturer's guidelines.

Sequence data analysis was carried out. Here, the Q25 sequence data derived from the sequencing process was processed using standard analysis pipeline processes (MR DNA, Shallowater, USA). Sequences were depleted of barcodes and primers then short sequences, 200 bp removed, as were sequences with ambiguous base calls removed, and sequences with homopolymer runs exceeding 6 bp, sequences were denoised and chimeras removed.[2-8] Operational taxonomic units were defined after removal of singleton sequences, clustering at 3% divergence (97% similarity). Final OTUs were taxonomically classified using BLASTn against a curated databased derived fromGreenGenes, NCBI and RDP databases.[9] Normalized and

de-noised files were then rarefied and run through QIIME[10] to generate alpha and beta diversity data. Additional statistical analyses were performed with NCSS2007 (NCSS, UT) and XLstat 2012 (Addinsoft, NY).

References for supplementary methods

1. Rogers GB, Kozlowska J, Keeble J, *et al.* Functional divergence in gastrointestinal microbiota in physically-separated genetically identical mice. Sci Rep. 2014:4;5437.

2. Dowd SE, Callaway TR, Wolcott RD, *et al.* Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol. 2008:8;125.

3. Dowd SE, Sun Y, Wolcott RD, *et al.* Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) for microbiome studies: bacterial diversity in the ileum of newly weaned Salmonella-infected pigs. Foodborne Pathog. Dis. 2008:5;459–72.

4. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010:26;2460–1.

5. Capone KA, Dowd SE, Stamatas GN, *et al.* Diversity of the human skin microbiome early in life. J. Invest. Dermatol. 2011:131;2026–32.

6. Dowd SE, Delton Hanson J, Rees E, *et al.* Survey of fungi and yeast in polymicrobial infections in chronic wounds. J. Wound Care 2011:20;40–7.

7. Eren AM, Zozaya M, Taylor CM, *et al.* Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. PLoS One 2011:6;e26732;

8. Swanson KS, Dowd SE, Suchodolski JS, *et al.* Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. ISME J. 2011:5;639–49.

9. DeSantis TZ, Hugenholtz P, Larsen N, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 2006:72;5069–72.

639 10. Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-throughput community
640 sequencing data. Nat Methods. 2010;7:335-6.